

0-798003

На правах рукописи

ГРИГОРЬЕВ Вениамин Юрьевич

**ДЕСКРИПТОРЫ ВОДОРОДНОЙ СВЯЗИ И
КОЛИЧЕСТВЕННЫЕ МОДЕЛИ «СТРУКТУРА–СВОЙСТВО»
ОРГАНИЧЕСКИХ СОЕДИНЕНИЙ**

02.00.04 – физическая химия

Автореферат
диссертации на соискание ученой степени
доктора химических наук

Черноголовка – 2012



Работа выполнена в Федеральном государственном бюджетном учреждении науки
Институте физиологически активных веществ РАН

Научный консультант: доктор химических наук, профессор Раевский Олег
Алексеевич

Официальные оппоненты: Касаикина Ольга Тарасовна, доктор химических наук,
профессор, Федеральное государственное бюджетное
учреждение науки Институт химической физики им.
Н.Н. Семенова РАН, г. Москва, зав. лабораторией
жидкофазного окисления

Дьячков Павел Николаевич, доктор химических наук,
профессор, Федеральное государственное бюджетное
учреждение науки Институт общей и неорганической
химии РАН, г. Москва, ведущий научный сотрудник
лаборатории квантовой химии

Поройков Владимир Васильевич, доктор
биологических наук, профессор, Федеральное
государственное бюджетное учреждение Научно-
исследовательский институт биомедицинской химии
им. В.Н. Ореховича РАН, г. Москва, зав. отделом
биоинформатики

Ведущая организация: Федеральное государственное образовательное
учреждение высшего профессионального образования
Московский государственный университет им. М.В.
Ломоносова, г. Москва

Защита состоится 24 октября 2012 г. в 10:00 на заседании диссертационного
совета Д 002.082.02 при Институте проблем химической физики РАН по адресу:
142432, Московская обл., г. Черноголовка, проспект академика Семенова, д. 1,
корпус общего назначения ИПХФ РАН.

С диссертацией можно ознакомиться в библиотеке ИПХФ РАН.

Автореферат разослан 12 сентября 2012 г.

Ученый секретарь диссертационного совета,
доктор химических наук



Джабиев Т.С.

НАУЧНАЯ БИБЛИОТЕКА КФУ



0000796121

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. Явление «водородной связи» известно уже более девяноста лет и интерес исследователей к нему никогда не пропадал ввиду многообразия проявлений Н-связи и той важности, которую она играет в природе. Водородная связь изучается с использованием различных физико-химических методов, среди которых термодинамические методы занимают одно из первых мест. Исследование термодинамики водородной связи между молекулами протонодонора и протонакцептора позволяет ответить на вопросы, касающиеся ее прочности (энтальпия), количества образующихся молекул Н-комплексов (энергия Гиббса или константа равновесия) и дает информацию о структурных изменениях (энтропия).

К настоящему времени в мире синтезировано около 20 миллионов химических соединений, большинство из которых относится к органическим веществам способным к образованию Н-связи. В сферу интересов человека вовлечено около 60 тысяч органических соединений, которые широко применяются в быту, медицине, на производстве и в сельском хозяйстве. Для многих из используемых соединений отсутствуют экспериментальные данные о некоторых их свойствах, например, об их биологической активности. Среди различных видов активности особое место занимает острая токсичность соединений *in vivo*, знание количественных характеристик которой необходимо для оценки опасности веществ по отношению к здоровью человека и к состоянию окружающей среды. Для определения величин LC_{50} и LD_{50} необходимы большие материальные и временные затраты, кроме того, это связано с использованием большого количества живых организмов.

Оценка количественных характеристик острой токсичности может осуществляться с использованием экспериментальных методов и математических моделей. При создании статистических моделей токсичности хорошие результаты были получены на основе междисциплинарного подхода «количественная связь структура – активность» (QSAR)/«количественная связь структура – свойство» (QSPR) При этом актуальными, требующими

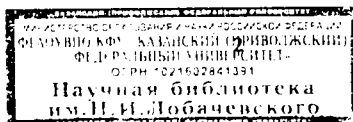
разрешения, являются вопросы, связанные с количественным учетом влияния водородной связи и разработкой соответствующих дескрипторов (факторов Н-связи).

Работа выполнена в соответствии с планами научно-исследовательских работ Института физиологически активных веществ РАН «Компьютерный молекулярный дизайн, включая количественную связь структура-свойство (активность) и молекулярное моделирование взаимодействия лиганд-мишень для структурной генерации соединений с заданными свойствами» (№ госрегистрации 0120.0 504023) при финансовой поддержке Российского фонда фундаментальных исследований: проект № 97-03-33068-а «Молекулярное моделирование связи структура-активность с учетом трехмерности структур и их способности к образованию водородных связей» и Международного научно-технического центра: проект № 888 «Компьютерное предсказание химической экотоксичности» и проект № 3777 «Развитие и применение КССА моделей токсичности химических соединений по отношению к грызунам».

Цель и задачи работы. Основная цель – создание основ физико-химического подхода для оценки протонодонорной и протоноакцепторной способности соединений и разработка количественных моделей острой токсичности *in vivo* органических соединений с использованием факторов водородной связи и других молекулярных дескрипторов.

Для достижения намеченной цели необходимо было решить следующие задачи:

- 1) разработать методы и дескрипторы для количественной оценки протонодонорной и протоноакцепторной способности органических соединений;
- 2) создать базу данных по термодинамическим характеристикам образования Н-комплексов и базу данных по факторам Н-связи;
- 3) разработать модель липофильности органических соединений;
- 4) создать и протестировать локальные и глобальные модели острой токсичности органических соединений;



- 5) разработать фрактальный дескриптор и модифицированную линейную динамическую модель для установления количественной связи между структурой и свойствами химических соединений;
- 6) сконструировать алгоритмы QSAR/QSPR моделирования и реализовать их в виде компьютерных программ.

Научная новизна. Разработан подход для расчета дескрипторов (факторов) водородной связи органических соединений, заключающийся в использовании аддитивно-мультипликативной модели и термодинамических характеристик Н-комплексов.

Предложена модель липофильности органических соединений, основанная на линейной комбинации двух молекулярных дескрипторов: поляризуемости и протонакцепторного фактора водородной связи.

Разработаны локальные модели неполярного и полярного наркоза, основанные на молекулярной поляризуемости и факторах Н-связи.

Созданы и применены алгоритмы для классификации химических соединений по типу токсического действия, отличающиеся от известных алгоритмов тем, что они используют локальные регрессионные модели наркоза.

Предложены глобальные модели острой токсичности, основанные на молекулярном сходстве, физико-химических дескрипторах и перекрывающихся кластерах.

Разработана рекуррентная модель для описания токсичности органических соединений, отличающаяся от известных моделей отсутствием дескрипторов.

Разработана модель ретардантной активности солей четвертичного аммония на основе факторов водородной связи и других физико-химических дескрипторов.

Впервые предложен и разработан фрактальный 3D дескриптор молекулярной структуры, представляющий из себя фрактальную размерность гистограмм межатомных расстояний.

Впервые разработан и применен подход для установления количественной связи между структурой и свойством (активностью), выраженными в виде двумерных кривых.

Практическая значимость работы. Значительная часть диссертационной работы была выполнена в рамках участия в разработке в ИФАВ РАН двух баз данных по Н-комплексобразованию органических соединений. При этом база данных по термодинамическим характеристикам образования Н-комплексов, насчитывающая свыше 13000 записей, нашла использование в качестве источника информации при проведении теоретических и практических исследований в области физической органической химии. База данных по протонодонорным и протоноакцепторным факторам Н-связи, содержащая свыше 50000 записей, а также компьютерная программа HUBOT, созданная на ее основе, служит для количественной оценки способности органических соединений к образованию водородной связи и нашла применение при проведении физико-химических расчетов. Разработанные в диссертации модели острой токсичности могут использоваться для количественной оценки токсического эффекта новых неисследованных органических соединений. Созданные компьютерные программы являются инструментом для проведения QSAR/QSPR исследований.

Личный вклад автора. Общее направление работы, включая постановку задач исследования, было определено совместно с научным консультантом. Все результаты диссертации получены лично автором или в соавторстве при его непосредственном участии. В работах, выполненных в соавторстве, вклад автора состоял в непосредственном участии во всех этапах работы начиная от постановки задачи и завершая обсуждением результатов. Синтез, очистку и идентификацию новых физиологически активных соединений выполнили сотрудники ИФАВ РАН: Иванов А.Н., Соколов В.Б., Чепакова Л.А., Брель В.К., Додонов М.В., Прошин А.Н., Чистяков В.Г. Биологические испытания на ретардантную активность провел Кокурин А.В. (ВНИИСБ РАСХН). Вклад

автора в программный комплекс НУВОТ заключался в создании базы факторов и термодинамической базы данных (совместно с Раевской О.Е., ИФАВ РАН).

Автор выражает глубокую признательность своему научному консультанту доктору химических наук, профессору Раевскому О.А.

Апробация работы. Результаты исследования и основные положения работы были представлены на следующих конференциях, симпозиумах и конгрессах: 11 Международная конференция по химии фосфора (Таллинн, 1989); IX Всесоюзная конференция «Химическая информатика» (Черноголовка, 1992); 12 Международный симпозиум по медицинской химии (Базель, Швейцария, 1992); 9 Европейский симпозиум по количественным соотношениям структура-активность (Страсбург, Франция, 1992); IV Российский конгресс «Человек и лекарство» (Москва, 1997); VI Российский конгресс «Человек и лекарство» (Москва, 1999); 4 Международный симпозиум по компьютерным методам в токсикологии и фармакологии (Москва, 2007); XVII Российский конгресс «Человек и лекарство» (Москва, 2010); 18 Европейский симпозиум по количественным соотношениям структура-активность (Родес, Греция, 2010); XVIII Российский конгресс «Человек и лекарство» (Москва, 2011), XIX Российский конгресс «Человек и лекарство» (Москва, 2012).

Публикации. По теме диссертации опубликовано 46 печатных работ, включая 1 учебное пособие, 24 статьи в российских и международных журналах, получено 2 свидетельства об официальной регистрации программ для ЭВМ.

Объем и структура диссертации. Работа изложена на 328 страницах машинописного текста, содержит 39 таблиц и 33 рисунка. Диссертация состоит из введения, пяти глав, заключения, списка литературы из 433 наименований.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность работы, сформулированы цель и задачи исследования, показана научная новизна и практическая значимость полученных результатов.

Глава 1. Современное состояние в области QSAR моделирования острой токсичности *in vivo* органических соединений. (Обзор литературы). Из анализа литературы следует, что во многих публикациях отмечается важная роль водородной связи при проявлении токсических эффектов. Однако, при этом способность соединений к образованию Н-связи выражается на полуколичественном уровне в виде числа доноров и акцепторов, без учета молекулярного окружения. По-видимому, качество моделей токсичности может быть значительно улучшено при применении адекватных количественных характеристик протондонорной и протонакцепторной способности органических соединений.

Наиболее часто используемым дескриптором для конструирования моделей острой токсичности органических соединений, относящихся по типу токсического действия к наркозу, является коэффициент распределения вещества в системе н-октанол-вода (K_{ow}). Однако, применение этого дескриптора приводит к появлению вопросов, связанных с объяснением различий между типами наркоза, главным образом, между неполярным и полярным наркозом. Одним из путей для решения этой проблемы может служить замена K_{ow} другими дескрипторами.

Число глобальных моделей токсичности, играющих важную роль при быстрой предварительной оценке токсичности большого числа разнообразных органических соединений, мало по сравнению с большим числом локальных моделей. Главным образом, это относится к изучению токсичности соединений по отношению к мышам и крысам.

Также можно сделать вывод о недостаточном числе исследований в области развития методологии QSAR. В первую очередь, это относится к разработке новых дескрипторов и способов установления связи между структурой и свойствами (активностью).

Глава 2. Экспериментальная часть. Описаны физико-химические и биологические данные; расчет, преобразование и отбор дескрипторов; меры сходства химических соединений; статистические методы исследования и

валидация моделей; методика проведения спектральных и калориметрических измерений.

Глава 3. Количественные характеристики протонодонорной и протоноакцепторной способности органических соединений

Энтальпийные и свободноэнергетические факторы водородной связи монофункциональных соединений

Для расчета факторов (дескрипторов) водородной связи органических соединений были использованы литературные данные, содержащие сведения о Н-комплексах и удовлетворяющие следующим условиям: 1) Н-комплекс состава 1:1, с одним водородным мостиком; 2) растворитель CCl_4 ; 3) для каждой реакции имеются значения ΔH (кДж/моль) и ΔG (кДж/моль); 4) ΔH и ΔG рассчитаны на основе экспериментальных данных и соответствуют температуре 298 К. Общее число Н-комплексов, образованных с участием 163 Н-доноров (NH-, OH-, SH-, CH-доноры) и 195 Н-акцепторов (N-, O-, S-, π -акцепторы), составило 936. Расчет факторов был выполнен с использованием аддитивно-мультипликативной модели:

$$\Delta H = k_1 E_a E_d + k_0 \quad (1)$$

$$\Delta G = k_1' C_a C_d + k_0' \quad (2)$$

где k_0 , k_1 , k_0' , k_1' – параметры; E_a – энтальпийный протоноакцепторный фактор; E_d – энтальпийный протонодонорный фактор; C_a – свободноэнергетический протоноакцепторный фактор; C_d – свободноэнергетический протонодонорный фактор. Для расчета E и C факторов необходима фиксация этих величин для одного Н-донора и одного Н-акцептора. Эти соединения должны удовлетворять двум требованиям: 1) для них должно быть описано достаточно большое количество Н-комплексов и 2) они должны быть удобными при экспериментальном применении. В качестве стандартного Н-донора был выбран фенол ($E_d = -2.50$; $C_d = -2.50$), в качестве стандартного Н-акцептора – гексаметилфосфортриамид (ГМФТА) ($E_a = 2.50$; $C_a = 4.00$). Величины факторов для стандартных соединений были выбраны таким образом, чтобы сделать шкалу с интервалом изменения факторов от 0.00 до ($\cong -5.00$) для Н-доноров и от

0.00 до (\cong 5.00) для Н-акцепторов. Алгоритм расчета протонодонорных факторов включал в себя следующие шаги: 1) найти протониакцепторы, которые имеют реакции с фенолом; 2) найти протонодоноры, которые имеют реакции с этими Н-акцепторами; 3) сконструировать и решить систему уравнений для каждой пары реакций, образованной фенолом и данным Н-донором с тем же самым протониакцептором. Протониакцепторные факторы рассчитывались с использованием того же самого алгоритма, используя ГМФТА в качестве стандартного Н-акцептора.

В результате удалось рассчитать величины факторов для 347 соединений. При этом величины E_d варьировали от -0.34 (C_6HCl_5) до -3.93 (CCl_3COOH), C_d от -0.00 (C_6HCl_5) до -4.78 (CCl_3COOH), E_a от 0.50 (C_6H_6) до 5.21 ($(C_4H_9)_2NH$), C_a от 0.43 (CCl_3CN) до 4.99 (1,4-диазобидикло[2,2,2]октан). Путем варьирования были определены коэффициенты моделей (1) и (2) и получены уравнения:

$$\Delta H(\text{кДж/моль}) = 4.96 E_a E_d \quad (3)$$

$$\Delta G(\text{кДж/моль}) = 2.43 C_a C_d + 5.70 \quad (4)$$

С учетом рассчитанных коэффициентов и используя полученные нами величины факторов для 163 протонодоноров и 195 протониакцепторов, можно оценить значения энтальпии и свободной энергии Гиббса для $163 \times 195 = 31785$ реакций на основе уравнений (3) и (4). Для имеющихся в нашем распоряжении 936 экспериментальных и рассчитанных величин ΔH и ΔG были получены следующие соотношения:

$$\Delta H_{\text{расч}} = -0.27(\pm 0.45) + 1.00(\pm 0.02) \Delta H_{\text{эксп}} \quad (5)$$

$$n=936; r=0.954; s=2.70; F=9553$$

$$\Delta G_{\text{расч}} = -0.07(\pm 0.12) + 1.00(\pm 0.01) \Delta G_{\text{эксп}} \quad (6)$$

$$n=936; r=0.984; s=1.11; F=28556$$

Статистические характеристики уравнений (5) и (6) свидетельствуют о том, что рассчитанные факторы дают возможность реальной оценки энтальпии и свободной энергии Гиббса Н-связи. При этом анализ остатков показал, что 842 из 936 рассчитанных величин ΔG имеют менее чем 10% отклонение от экспериментальных величин.

В литературе, кроме данных по образованию Н-комплексов в CCl_4 , приводятся величины ΔH и ΔG в других неполярных растворителях. С этими данными была проведена работа аналогичная рассмотренной выше. Из анализа полученных результатов можно сделать вывод о том, что факторы соединений сохраняют приблизительно одинаковое значение в разных растворителях, при этом происходит только изменение коэффициентов в уравнениях (1) и (2).

Энтальпийные и свободноэнергетические факторы водородной связи полифункциональных соединений

Для проверки работоспособности аддитивно-мультипликативной модели и факторов в случае сложных органических соединений были проведены экспериментальные исследования с использованием ИК-спектроскопии и калориметрии. Они включали в себя установление стехиометрии, определение взаимодействующих атомов и групп, расчет величин ΔH , ΔG для Н-комплексов и оценку величины факторов. В исследовании принимали участие 22 фосфорилированных и тиофосфорилированных оксима общей формулы: $(R_1)(R_2)P(X)ONC(R_3)(R_4)$ (где $R_1=\text{Me, MeO, EtO, PrO, BuO}$; $R_2=\text{MeO, EtO, PrO, i-PrO, BuO, AmO}$; $R_3=\text{Me, Et, Pr, CH}_2\text{Cl}$; $R_4=\text{Me, Et, Pr, Cl}$; $X=\text{O, S}$), которые можно рассматривать в качестве аналогов инсектицида «валексон» $(\text{EtO})_2\text{P}(\text{S})\text{ONC}(\text{C}_6\text{H}_5)\text{CN}$ и 10 дифторнитроацетанилидов общей формулы: $(4-R_5)(3-R_6)\text{C}_6\text{H}_3\text{NHC}(\text{O})\text{CF}_2\text{NO}_2$ (где $R_5=\text{H, CH}_3, \text{OCH}_3, \text{F, Cl, NO}_2$; $R_6=\text{H, CH}_3, \text{CF}_3, \text{F, NO}_2$), которые можно рассматривать в качестве аналогов гербицида «пропанид» $3,4\text{-Cl}_2\text{-C}_6\text{H}_3\text{NHC}(\text{O})\text{C}_2\text{H}_5$.

При изучении ИК спектров растворов комплексов фосфорилированных оксимов с фенолом и NH-донорами наблюдается картина типичная при образовании водородной связи. На рис. 1 изображен ИК спектр системы фенол + соединение $(\text{Me})(\text{EtO})\text{P}(\text{O})\text{ONC}(\text{Me})(\text{Cl})$ (I) в гексане. Узкая полоса при 3624 см^{-1} соответствует валентным колебаниям гидроксильной группы мономерных молекул фенола, а широкая полоса в области $3315\text{-}3325\text{ см}^{-1}$ соответствует валентным колебаниям гидроксильных групп фенола, связанного в комплекс с соединением (I). Наряду со сдвигом полосы $\nu(\text{OH})$ в область низких частот

наблюдается усиление интенсивности этой полосы. Так, молярный коэффициент экстинкции ϵ полосы $\nu(\text{OH})$ свободных от ассоциации молекул фенола составляет величину 202 л/(моль см), а связанных в комплекс – 387 л/(моль см), т.е. ϵ возрастает в $(387/202)=1.92$ раза. Исходя из структурной формулы фосфорилированных оксимов можно заключить, что, очевидно, наиболее сильным Н-акцепторным центром в молекулах этих соединений является фосфорильный атом кислорода. Это подтверждает изученный ИК-спектр комплекса соединения (I) с фенолом в области валентных колебаний фосфорильной группы (рис. 1). Полоса при 1275 см^{-1} отвечает валентным колебаниям $\nu(\text{PO})$. При комплексообразовании происходит сдвиг этой полосы на $30\pm 50\text{ см}^{-1}$ в область низких частот, что свидетельствует об участии фосфорильной группы в Н-связывании. На основе полученных данных были рассчитаны термодинамические характеристики образования Н-комплексов состава 1:1, которые менялись от -18.5 до -28.3 кДж/моль для ΔH и от -6.05 до -16.3 кДж/моль для ΔG .

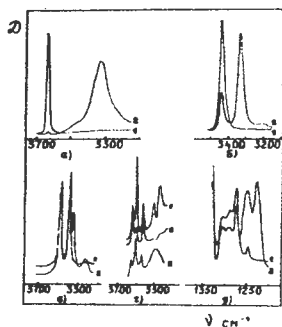


Рис. 1. ИК-спектры растворов соединений в гексане. а) ИК-спектры растворов фенола в области $\nu(\text{OH})$. 1) 0.0219 моль/л PhOH; 2) 0.0110 моль/л PhOH + 0.0250 моль/л $(\text{Me})(\text{EtO})\text{P}(\text{O})\text{ONC}(\text{Me})(\text{Cl})$ (I). б) ИК-спектры растворов дифениламина в области $\nu(\text{NH})$. 1) 0.0944 моль/л Ph_2NH ; 2) 0.0472 моль/л Ph_2NH + 0.0250 моль/л (I). в) ИК-спектры растворов анилина в области $\nu(\text{NH})$. 1) 0.0769 моль/л PhNH_2 ; 2) 0.0769 моль/л PhNH_2 + 0.0250 моль/л (I). г) ИК-спектры растворов уретана в области $\nu(\text{NH})$. 1) 0.0177 моль/л $\text{C}_2\text{H}_5\text{OC}(\text{O})\text{NH}_2$ (ус. X1); 2) 0.0018 моль/л $\text{C}_2\text{H}_5\text{OC}(\text{O})\text{NH}_2$ (ус. X5); 3) 0.0018 моль/л $\text{C}_2\text{H}_5\text{OC}(\text{O})\text{NH}_2$ + 0.0250 моль/л (I) (ус. X5); д) ИК-спектры растворов соединения (I) в области $\nu(\text{PO})$. 1) 0.0054 моль/л (I); 2) 0.0054 моль/л (I) + 0.0107 моль/л PhOH

При взаимодействии дифторнитроацетанилидов с ГМФТА образуются Н-комплексы состава 1:1, в которых группа NH выступает в качестве Н-донора, а фосфорильный кислород - в качестве Н-акцептора. ΔH меняется в пределах от -24.9 до -30.5 кДж/моль, а ΔG от -15.0 до -20.8 кДж/моль. С использованием полученных экспериментальных величин были рассчитаны факторы фосфорилированных оксимов и дифторнитроацетанилидов.

Следующим шагом стало изучение комплексообразования с участием этих соединений, в котором фосфорилированные оксими выступали в качестве Н-акцепторов, а дифторнитроацетанилиды – в качестве Н-доноров (табл. 1).

Таблица 1

Экспериментальные и рассчитанные термодинамические характеристики (ΔH , ΔG) Н-комплексов фосфорилированных оксимов с дифторнитроацетанилидами (бензол; 298 К; состав 1:1) и факторы Н-связи (E_a , C_a , E_d , C_d)

№	Н-акцептор		№	Н-донор		- ΔH , кДж/моль		- ΔG , кДж/моль	
	E_a	C_a		E_d	C_d	Эксп.	Расч.*	Эксп.	Расч.*
II	1.71±0.09	3.08±0.11	V	2.71±0.06	2.41±0.07	18.9±1.0	20.2±1.5	10.5±0.6	10.4±1.0
II	1.71±0.09	3.08±0.11	V	2.71±0.06	2.41±0.07	18.9±1.0	20.2±1.5	10.5±0.6	10.4±1.0
II	1.71±0.09	3.08±0.11	VI	2.48±0.02	3.01±0.06	18.9±1.0	18.5±1.1	13.8±0.6	14.4±1.1
III	1.73±0.09	2.84±0.11	V	2.71±0.06	2.41±0.07	19.4±1.0	20.4±1.5	9.98±0.6	9.24±1.0
III	1.73±0.09	2.84±0.11	VI	2.48±0.02	3.01±0.06	18.9±1.0	18.7±1.1	13.5±0.6	12.4±1.1
IV	1.75±0.09	3.04±0.11	V	2.71±0.06	2.41±0.07	19.8±1.0	20.6±1.5	11.0±0.6	10.3±1.0
IV	1.75±0.09	3.04±0.11	VI	2.48±0.02	3.01±0.06	19.6±1.0	18.9±1.1	15.3±0.6	14.2±1.1

Примечание: $\Delta H = 4.35 E_a E_d$; $\Delta G = 5.70 + 2.17 C_a C_d$; (Me)(BuO)P(O)ONC(Me)(Cl) (II); (Me)(BuO)P(O)ONC(Et)(Cl) (III); (Me)(BuO)P(O)ONC(Pr)(Cl) (IV); 4-CH₃-C₆H₄NHC(O)CF₃NO₂ (V); 4-NO₂-C₆H₄NHC(O)CF₃NO₂ (VI).

При сравнении экспериментальных и рассчитанных на основе факторов Н-связи термодинамических величин можно установить, что в случае ΔH минимальное отклонение (по абсолютной величине) составляет 0.2 кДж/моль, максимальное 1.3 кДж/моль и среднее 0.6 кДж/моль, а для ΔG 0.1 кДж/моль, 1.1 кДж/моль и 0.8 кДж/моль соответственно. Таким образом, можно констатировать вполне удовлетворительное совпадение экспериментальных и рассчитанных термодинамических характеристик образования Н-комплексов.

QSPR модель липофильности органических соединений

Для количественного описания липофильности (K_{ow} , коэффициент распределения вещества в системе октанол-вода) было отобрано 71 простое

органическое соединение, содержащее по одной Н-акцепторной группе. В состав этих групп входили атомы О, N, S, F, Cl, Br, С. В качестве независимых переменных были использованы молекулярная поляризуемость (α) и свободноэнергетические Н-акцепторные факторы (C_a). Интервал изменения этих величин составлял $3.25 \pm 21.44 \text{ \AA}^3$ и 0.20 ± 3.12 соответственно. Величины $\log K_{ow}$ менялись в пределах от -0.97 до 4.75. В результате была получена следующая модель:

$$\log K_{ow} = 0.266(\pm 0.006) \alpha - 1.00(\pm 0.05) C_a \quad (7)$$

$n=71; r^2=0.982; s=0.18; F=3829; q^2=0.981$

Статистические критерии уравнения (7) свидетельствуют о возможности количественной оценки значений $\log K_{ow}$ на основе рассчитанных величин α и C_a . Внешняя валидация модели была проведена путем расчета аналогичной модели с использованием другого ряда из 98 соединений, содержащих протоноакцепторные атомы О, N, Cl, Br, I, С. Интервал изменения величин составлял -1.01 ± 5.15 , $2.25 \pm 28.77 \text{ \AA}^3$, 0.00 ± 2.96 для $\log K_{ow}$, α , C_a соответственно. Полученные параметры оказались близки к соответствующим параметрам модели (7):

$$\log K_{ow} = 0.276(\pm 0.008) \alpha - 1.01(\pm 0.07) C_a \quad (8)$$

$n=98; r^2=0.961; s=0.29; F=1191; q^2=0.959$

Преобразование уравнения (7) к следующему виду:

$$C_a^* = 0.266 \alpha - \log K_{ow} \quad (9)$$

дает возможность рассчитать свободноэнергетические протоноакцепторные факторы (C_a^*) соединений, основываясь на экспериментальных данных по их липофильности.

Имеется много опубликованных данных по экспериментальным величинам $\log K_{ow}$ как простых, так и сложных нейтральных и ионизированных соединений. И это дает возможность на основе уравнения (9) оценить протоноакцепторную способность тех атомов и функциональных групп, которые не могли быть определены в рамках шкалы, основанной на термодинамических параметрах (табл. 2).

Таблица 2

Свободноэнергетические Н-акцепторные факторы, полученные на основе липофильности (C_a^*), свободной энергии (C_a) и их разность (ΔC_a) для органических соединений, содержащих функциональные группы с несколькими гетероатомами

№	Соединение	C_a^*	C_a	ΔC_a
1	м-Нитробензойная кислота	0.62	0.47	0.15
2	Бензойная кислота	1.03	0.76	0.27
3	Этилтрихлорацетат	1.21	1.11	0.10
4	Уксусная кислота	1.47	1.10	0.37
5	Метилацетат	1.60	1.40	0.20
6	Бутиролактон	2.71	2.05	0.66
7	N,N-Диметилацетамид	3.02	2.76	0.26
8	Мочевина	3.06	2.50	0.56
9	N-Метилацетамид	3.15	2.70	0.45
10	2-Пиперидон	3.34	2.95	0.39

Представленные данные свидетельствуют о том, что в случае сложных органических соединений величина C_a^* превышает значение фактора для наиболее сильного Н-акцептора (C_a) и, вероятно, соответствует сумме свободноэнергетических протонакцепторных факторов в молекуле, в то время как их разность (ΔC_a) характеризует потенциальную возможность к образованию Н-связи слабых Н-акцепторных центров в молекуле. Это положение открывает широкие возможности для существенного увеличения числа соединений с количественно охарактеризованной протонакцепторной способностью.

Компьютерный программный комплекс НУВОТ

НУВОТ состоит из трех главных компонентов. Во-первых, это база данных (БД) по термодинамике водородной связи, которая содержит экспериментальные величины энтальпии и свободной энергии Гиббса образования Н-комплексов в различных средах. В настоящее время эта БД насчитывает 13688 записей (т.е. сведений об одной паре реакций Н-донор/Н-акцептор). БД содержит ряд полей: свободная энергия Гиббса образования Н-комплекса; энтальпия образования Н-комплекса; метод определения свободной

энергии Гиббса; метод определения энтальпии; разница в химических сдвигах между свободными и связанными Н-донорными группами; разница между частотами колебаний свободных и связанных Н-донорных групп; тип водородного мостика в комплексе АН...В, где АН – протонодонорная группа, В – протоноакцепторный атом; температура, к которой относятся ΔH и ΔG ; растворитель; стехиометрия Н-комплекса; число водородных мостиков в Н-комплексе; комментарии к ссылке и данным; авторы публикации; наименование источника данных; год публикации; том журнала; номер журнала; страницы публикации; молекулярная масса соединения; название химического соединения; номер CAS; название; брутто-формула соединения; химическая структура Н-донора; химическая структура Н-акцептора.

Другой составной частью НУВОТ является факторная база данных, содержащая протонодонорные и протоноакцепторные факторы органических соединений, рассчитанных в рамках общей шкалы водородной связи. Эта БД содержит 4 библиотеки факторов: E_d (1025 факторов), E_a (2060 факторов), C_d (2835 факторов), C_a (46372 фактора). БД содержит ряд полей: химическая структура соединения; название соединения; брутто-формула соединения; молекулярная масса соединения; номер CAS; величина фактора; активный центр.

Третьей составной частью НУВОТ является программа, которая рассчитывает факторы и дескрипторы органических соединений на основе их химической структуры с использованием баз данных, содержащих термодинамические характеристики Н-комплексов и факторы Н-связи.

Глава 4. QSAR модели острой токсичности органических соединений

Классификационные и регрессионные модели неполярного наркоза

В работе были использованы данные по острой токсичности (LC_{50} , ммоль/л, 96 час.) 118 органических соединений с типом токсического действия (MOA, Mode of action) неполярный наркоз по отношению к рыбам *Poecilia reticulata* (Pr), *Pimephales promelas* (Pp) и *Oncorhynchus mykiss* (Om). В качестве независимых переменных при конструировании регрессионных моделей были

использованы экспериментальные величины $\log K_{ow}$ и три физико-химических дескриптора: α , ΣC_a , ΣC_d . Химические соединения относились к трем структурным областям: хлоралканы и хлорбензолы (слабые Н-акцепторы), эфиры и кетоны (средние Н-акцепторы) и соединения с ОН-группами. Интервал изменения величин составлял для $\log(1/LC_{50})$: -2.96 ± 3.15 , $\log K_{ow}$: -1.36 ± 5.18 , α : $3.25 \pm 23.43 \text{ \AA}^3$, ΣC_a : 0.18 ± 5.77 , ΣC_d : -4.16 ± 0.00 . Для проведения классификации рассчитывались дескрипторы соединений с использованием программ HYBOT (32 дескриптора), DNESTR (48 дескрипторов) и SYBYL (9 дескрипторов). Фрагментные дескрипторы были рассчитаны с помощью программы MOLDIVS. Классификация выполнена на основе линейного дискриминантного анализа (LDA) с применением ряда из 4-х HYBOT дескрипторов α , Σq , ΣC_a , ΣC_d (LDA_4D), ряда из 10-12 главных компонент, полученных из 89 дескрипторов (LDA_PC) и метода структурного сходства, (аналог метода k-ого ближайшего соседа) с использованием фрагментных дескрипторов и метрики в виде коэффициентов Танимото (SS).

По сравнению с другими используемыми подходами метод SS оказывается наиболее эффективным при дискриминации соединений с МОА неполярный наркоз и соединений с другими типами токсического действия (табл. 3). Действительно, общая степень правильного распознавания составляет 0.932 и 0.864 в обучающей выборке, 0.914 и 0.881 в тестовой выборке для данных по отношению к Pr и $(Pr + Pp + Om)$ соответственно.

Учитывая высокую степень межвидовой корреляции значений токсичности для трех рыб ($r^2=0.94 \pm 0.99$), вполне возможно построение общей регрессионной модели токсичности:

$$\log(1/LC_{50}) = -1.75(\pm 0.05) + 0.89(\pm 0.03) \log K_{ow} \quad (10)$$

$n=150$; $r^2=0.936$; $s=0.36$; $F=2171$

Уравнение (10) не является новым с позиции описания неполярного наркоза на основе экспериментальных величин $\log K_{ow}$. Однако, общая модель для трех видов рыб позволяет предсказывать величины острой токсичности для одного

из видов на основе экспериментальных величин токсичности для других видов, используя $\log K_{ow}$.

Таблица 3

Чувствительность (Sn), специфичность (Sp) и точность (Acc) при бинарной классификации химических соединений с МОА неполярный наркоз (класс 1) и другие МОА (класс 2) по отношению к рыбам *P. reticulata* (Pr), *P. promelas* (Pp) и *O. mykiss* (Om) с использованием методов линейного дискриминантного анализа (LDA) и структурного сходства (SS)

№	Объект	Метод	Обучающая выборка				Тестовая выборка			
			n_1/n_2	Sn	Sp	Acc	n_1/n_2	Sn	Sp	Acc
1	Pr	LDA_4D	72/163	0.472	0.859	0.740	18/40	0.333	0.850	0.690
2	Pr	LDA_PC	72/163	0.639	0.834	0.774	18/40	0.556	0.825	0.741
3	Pr	SS_3	72/163	0.986	0.908	0.932	18/40	0.944	0.900	0.914
4	Pr+Pp+Om	LDA_4D	95/244	0.263	0.877	0.705	23/61	0.261	0.934	0.750
5	Pr+Pp+Om	LDA_PC	95/244	0.505	0.857	0.758	23/61	0.348	0.934	0.774
6	Pr+Pp+Om	SS_1	95/244	0.842	0.873	0.864	23/61	0.913	0.869	0.881

Примечание: n_1 – число соединений в первом классе, n_2 – число соединений во втором классе.

Подобная процедура возможна для конструирования общей QSAR модели на основе молекулярной поляризуемости и Н-акцепторных факторов:

$$\log(1/LC_{50}) = -1.96(\pm 0.09) + 0.258(\pm 0.006) \alpha - 0.88(\pm 0.03) \sum C_a \quad (11)$$

$$n=150; r^2=0.943; s=0.34; F=1211$$

Статистические характеристики уравнения (11) несколько лучше, чем характеристики уравнения (10). Однако, главное преимущество уравнения (11) состоит в использовании не экспериментальных, а рассчитанных дескрипторов. Кроме того, уравнение (11) дает возможность количественной оценки вкладов с противоположными эффектами: молекулярной поляризуемости соединений, которая увеличивает токсичность, и Н-акцепторной способности, которая уменьшает ее. Включение в модель дополнительного дескриптора $\sum C_d$ не улучшало статистические характеристики, а соответствующий коэффициент в модели незначимо отличался от нуля. Валидация уравнений (10) - (11) была выполнена с использованием тестового ряда:

$$\log(1/LC_{50})_{\text{эксп}} = -0.05(\pm 0.06) + 1.02(\pm 0.05) \log(1/LC_{50})_{\text{расч}} \quad (12)$$

$$n=37; r^2=0.930; s=0.36; F=462$$

$$\log(1/LC_{50})_{\text{эксп}} = 0.00(\pm 0.05) + 0.99(\pm 0.06) \log(1/LC_{50})_{\text{расч}} \quad (13)$$

$$n=37; r^2=0.942; s=0.33; F=564$$

Полученные результаты свидетельствуют о хорошей описательной и предсказательной способности моделей неполярного наркоза в пределах области применимости, определяемой интервалами изменения соответствующих величин.

Классификационные и регрессионные модели полярного наркоза

Для исследования была использована выборка, содержащая данные по острой токсичности (LC_{50} , ммоль/л, 96 час.) 134 соединений с МОА полярный наркоз по отношению к трем рыбам (*Pr*, *Pp* и *Om*). Анализируемые соединения можно отнести к двум структурным областям: соединения без Н-донорной способности (нитробензолы) и соединения с Н-донорной способностью (амины, анилины, фенолы). Интервал изменения величин составлял для $\log(1/LC_{50})$: $-1.54 \div 3.58$, $\log K_{ow}$: $-1.49 \div 6.32$, α : $5.80 \div 30.26 \text{ \AA}^3$, ΣC_a : $0.48 \div 4.18$, ΣC_d : $-6.52 \div 0.00$.

Использование метода LDA_4D привело к получению удовлетворительной дискриминации между соединениями с МОА полярный наркоз и соединениями с другими МОА (табл. 4). Так, например, степень общего правильного распознавания в случае данных для *Pr* составила 0.855 и 0.860 в обучающей и тестовой выборках соответственно. Использование для дискриминации соединений метода SS на основе первого структурного соседа (SS_1) приводило к получению хороших результатов. Например, применение SS_1 для обучающего ряда в тестовом ряду соединений по отношению к *Pr* обеспечивало общую степень правильного распознавания равную 0.894.

Высокий коэффициент корреляции межвидовой токсичности между рыбами *P. reticulata* и *P. promelas* ($r^2=0.91$) свидетельствует о возможности создания общей регрессионной модели:

$$\log(1/LC_{50}) = -0.78(\pm 0.06) + 0.72(\pm 0.02) \log K_{ow} \quad (14)$$

$$n=169; r^2=0.860; s=0.37; F=1022; q^2=0.857$$

$$\log(1/LC_{50}) = -1.02(\pm 0.14) + 0.192(\pm 0.007) \alpha - 0.68(\pm 0.04) \sum C_a - 0.11(\pm 0.02) \sum C_d \quad (15)$$

$$n=169; r^2=0.886; s=0.34; F=427; q^2=0.879$$

Таблица 4

Чувствительность (Sn), специфичность (Sp) и точность (Acc) при бинарной классификации химических соединений с МОА полярный наркоз (класс 1) и другие МОА (класс 2) по отношению к рыбам *P. reticulata* (Pr) и *P. promelas* (Pp) с использованием методов линейного дискриминантного анализа (LDA) и структурного сходства (SS)

№	Объект	Метод	Обучающая выборка				Тестовая выборка			
			n_1/n_2	Sn	Sp	Acc	n_1/n_2	Sn	Sp	Acc
1	Pr	LDA_4D	96/138	0.792	0.899	0.855	23/34	0.826	0.882	0.860
2	Pr	LDA_PC	96/138	0.833	0.899	0.872	23/34	0.870	0.882	0.877
3	Pr	SS_1	96/138	0.896	0.913	0.906	23/34	0.957	0.941	0.947
5	Pr+Pp	LDA_4D	106/280	0.623	0.925	0.842	26/70	0.654	0.871	0.812
6	Pr+Pp	LDA_PC	106/280	0.594	0.921	0.832	26/70	0.538	0.857	0.771
7	Pr+Pp	SS_1	106/280	0.868	0.904	0.894	26/70	0.923	0.914	0.917

Примечание: n_1 – число соединений в первом классе, n_2 – число соединений во втором классе.

Необходимо отметить, что дескриптор $\log K_{ow}$ не дает информации о механизме действия, потому что он является композитным дескриптором, включающим в себя несколько физико-химических дескрипторов. С другой стороны, модели, использующие HVBOT дескрипторы для описания неполярного и полярного наркоза, позволяют заглянуть внутрь механизма неспецифической токсичности. Целесообразно сравнить соответствующие модели по отношению к *P. reticulata* + *P. promelas* для полярного наркоза (уравнение (15)) и неполярного наркоза (уравнение (16)).

$$\log(1/LC_{50}) = -2.01(\pm 0.09) + 0.256(\pm 0.006) \alpha - 0.84(\pm 0.02) \sum C_a \quad (16)$$

$$n=156; r^2=0.946; s=0.32; F=1336; q^2=0.943$$

«Чистый» эффект полярного наркоза (полярный наркоз – неполярный наркоз):

$$\Delta \log(1/LC_{50}) = 0.99(\pm 0.23) - 0.064(\pm 0.013) \alpha + 0.16(\pm 0.06) \sum C_a - 0.11(\pm 0.02) \sum C_d \quad (17)$$

Из представленного уравнения ясно, что рост полярности молекул ведет к уменьшению разницы токсичностей, вычисленных по модели полярного и неполярного наркоза, в то время как рост числа H-акцепторных и H-донорных

центров в молекуле приводит к увеличению этой разницы. Полученный результат вместе с заметным вкладом протонодонорного фактора $\sum C_a$ в полярный наркоз, свидетельствует о существовании различия между соединениями, относящимися по типу токсического действия к неполярному и полярному наркозу и обусловленному дополнительным межмолекулярным взаимодействием с клеточными мембранами.

Модель АЛБТ

Исследование было проведено на основе ряда из 9618 разнообразных органических соединений, обладающих острой внутривенной токсичностью по отношению к мышам (LD_{50} , ммоль/кг, 24 час.). Интервал изменения величины $\log(1/LD_{50})$ составлял $-2.23 \div 6.73$. α и $\sum C_a$ менялись в пределах от 3.25 до 132 \AA^3 и от 0.00 до 30.6 соответственно.

Алгоритм классификации с использованием анализа локальной базовой токсичности (АЛБТ) может быть описан в виде следующих шагов: расчет величин молекулярной поляризуемости α , суммы свободнoэнергетических протонoакцепторных факторов $\sum C_a$ и коэффициентов Танимото T_c для исходной выборки молекул; деление исходной выборки на обучающую и тестовую; задание шага изменения величин $\Delta \sum C_a$, $\Delta \alpha$ и токсичности $\Delta \log(1/LD_{50})$; кластеризация молекул обучающей выборки с применением в качестве ведущей характеристики $\sum C_a$ (генерация интервалов с шагом $\Delta \sum C_a$ от 0 до $\max \sum C_a$); разбиение каждого молекулярного кластера на ячейки с использованием α (интервалы с шагом $\Delta \alpha$ от 0 до $\max \alpha$); идентификация в каждой ячейке молекул с минимальной величиной $\log(1/LD_{50})$; создание для каждого кластера ряда молекул с минимальными величинами токсичности; расчет линейных или параболических регрессионных зависимостей $\log(1/LD_{50})$ от α для созданного ряда молекул; расчет на основе полученных моделей величин острой токсичности для всех молекул кластера; классификация молекул кластера на две группы: (неполярный наркоз)/(все остальное) путем сравнения экспериментальных $(\log(1/LD_{50}))_{\text{эксп}}$ и рассчитанных $(\log(1/LD_{50}))_{\text{расч}}$ значений токсичности с учетом «коридора» токсичности (интервалы

$\log(1/LD_{50})_{расч} \pm \Delta \log(1/LD_{50})$ с шагом $\Delta \log(1/LD_{50})$ от 0 до $\max \Delta \log(1/LD_{50})$); отнесение каждой молекулы обучающей выборки к соответствующему кластеру обучающей выборки на основе величины $\sum C_i$; определение для каждой молекулы обучающей выборки ближайшего соседа с максимальной величиной T_c из соответствующего кластера; классификация молекул обучающей выборки на две группы: (неполярный наркоз)/(все остальное) на основе принадлежности ближайшего соседа к той или иной группе; расчет характеристик бинарной классификации молекул; сортировка характеристик бинарной классификации молекул; выбор лучшей модели; повторение шагов для молекул тестовой выборки.

В настоящей работе в качестве сравнительных методов классификации химических соединений использованы такие методы, как линейный дискриминантный анализ (LDA), k-ближайший сосед (KNN), структурное сходство (SS), опорные вектора (SVM), нейронная сеть (ANN). Классификационные методы требуют априорного знания о принадлежности соединений обучающей выборки к тому или иному классу. Так как для исследуемых 9618 органических соединений такая информация отсутствовала, то для их разделения на два класса (неполярный наркоз)/(все остальное) мы использовали метод аналогии, основываясь на данных по типам токсического действия органических соединений по отношению к водным организмам. Для этого были отобраны 42 классических инертных соединения с типом токсического действия неполярный наркоз, которые включали в себя спирты, кетоны, эфиры. Для них была сконструирована билинейная модель зависимости острой токсичности от расчетных величин коэффициента распределения вещества в системе н-октанол-вода (P):

$$\log(1/LD_{50}) = -1.54(\pm 0.08) + 0.82(\pm 0.05) \log P - 1.399(\pm 0.118) \log[0.001607 P + 1] \quad (18)$$

$$n = 42; r^2 = 0.868; s = 0.24; F = 129; q^2 = 0.852$$

Необходимо подчеркнуть, что в указанной выборке из 42 соединений представлены данные для мышей различного пола, возраста, условий содержания и эксперимента. Влияние этих факторов может давать различия в

значениях $\log(1/LD_{50})$ на уровне ± 0.50 . Поэтому к 42 соединениям с MOA неполярный наркоз мы добавили все соединения, которые имеют экспериментальные величины токсичности, попадающие в коридор (уравнение (18)) ± 0.50 . Вошедшие в базовый коридор 3568 соединений имеют значения $\log P$ в интервале $-0.81 \div 5.42$. Оставшиеся 6050 соединений расположены вне базового коридора: 5833 - над ним, 217 - под ним.

В табл. 5 представлены результаты бинарной классификации 9618 органических соединений с использованием различных методов в пространстве из 16 HYBOT дескрипторов (LDA, KNN, SVM, ANN), фрагментных дескрипторов (SS), α и ΣC_a дескрипторов (АЛБТ). Следует отметить, что среди использованных методов самые худшие результаты показывает метод LDA, а самые лучшие – АЛБТ. Так, в случае LDA, величина Асс, составляет 0.666 (обучающая выборка) и 0.652 (тестовая выборка). При использовании АЛБТ Асс достигает 0.998 и 0.824 соответственно. Подходы, основанные на анализе ближайшего соседа (KNN_7 и SS_5), показывают промежуточные результаты.

Таблица 5

Чувствительность (Sn), специфичность (Sp) и точность (Асс) при бинарной классификации химических соединений с MOA неполярный наркоз (класс 1) и другие MOA (класс 2) по отношению к мышам с использованием методов линейного дискриминантного анализа (LDA), k-ближайшего соседа (KNN), структурного сходства (SS), опорных векторов (SVM), нейронных сетей (ANN) и анализа локальной базовой токсичности (АЛБТ)

№	Метод	Обучающая выборка				Тестовая выборка			
		n_1/n_2	Sn	Sp	Асс	n_1/n_2	Sn	Sp	Асс
1	LDA	2811/4884	0.291	0.881	0.666	757/1166	0.284	0.890	0.652
2	KNN_7	2811/4884	0.605	0.807	0.733	757/1166	0.594	0.813	0.727
3	SS_5	2811/4884	0.587	0.798	0.721	757/1166	0.568	0.810	0.715
4	SVM	2811/4884	0.855	0.947	0.913	757/1166	0.567	0.852	0.739
5	ANN	2811/4884	0.406	0.853	0.690	757/1166	0.394	0.846	0.668
6	АЛБТ	1959/5715	0.992	1.000	0.998	509/1405	0.650	0.887	0.824

Примечание: n_1 – число соединений в первом классе, n_2 – число соединений во втором классе.

Модель АМТ

АМТ (арифметическая средняя токсичность) подход был использован для предсказания острой внутривенной токсичности (LD_{50} , ммоль/кг, 24 час.) 10241 органического соединения по отношению к мышам. Полный интервал величин $\log(1/LD_{50})$ составлял 7.49 единиц $(-3.14 \div 4.35)$. Для описания структуры соединений были рассчитаны 32 НУВОТ дескриптора и коэффициенты Танимото (T_c).

Алгоритм модели АМТ может быть описан в виде четырех шагов. Шаг 1. Для каждой анализируемой молекулы определялась величина структурного сходства с остальными молекулами. Полученные величины T_c упорядочивались по убыванию. Путем задания минимального граничного значения T_c^* формировались различного размера первичные кластеры ближайших структурных соседей. Шаг 2. Для всех молекул рассчитывался ряд дескрипторов. Шаг 3. С использованием каждого дескриптора в отдельности и первичного структурного кластера для каждой молекулы рассчитывался вторичный кластер. Для этого из молекул первичного кластера выбирались такие пары ближайших соседей, которые имели значение дескриптора больше и меньше, чем у анализируемой молекулы. Размер вторичного кластера регулировался путем задания числа молекулярных пар (NP). Шаг 4. Токсичность анализируемой молекулы рассчитывалась как арифметическое среднее токсичностей молекулярных пар. Наилучшие результаты были получены на основе НУВОТ дескриптора $\sum C_{ad}/\alpha$. Поэтому этот дескриптор был использован, чтобы получить результаты, описанные ниже.

Путем варьирования значений T_c^* (0.00, 0.40, 0.50, 0.60, 0.70, 0.80) и NP (1, 2, 3, 4, 5) был проведен расчет величин $\log(1/LD_{50})$. Качество полученных моделей оценивалось путем сравнения экспериментальных и рассчитанных величин токсичности с использованием уравнения:

$$\log(1/LD_{50})_{\text{эксп}} = a_0 + a_1 \log(1/LD_{50})_{\text{расч}} \quad (19)$$

Интервал изменения параметров и статистических характеристик полученных уравнений был следующим: a_0 : $-0.02 \div 0.15$; a_1 : $0.77 \div 1.10$; n: 491+10228; r^2 :

0.456±0.783; s: 0.33±0.52; F: 1567±9672; q^2 : 0.456±0.773. При возрастании границы структурного сходства и числа пар, свободный член уравнения (19) приближается к 0, а угловой коэффициент приближается к 1. В то же самое время, коэффициент линейной корреляции возрастает, а стандартное отклонение убывает. Однако, этот процесс сопровождается уменьшением числа соединений, вовлеченных в расчет.

В принципе, почти для всех химических соединений изучаемого ряда могут быть рассчитаны величины токсичности на основе модели АМТ с использованием $T \geq 0.00$ и $NP=2$. Только 14 соединений из 10241 не имеют двух пар соседей с большим и меньшим значением выбранного дескриптора ($\sum C_{ad}/\alpha$). Параметры уравнения (19) в этом случае удовлетворительные и стандартное отклонение сравнимо с ожидаемыми экспериментальными ошибками измерения токсичности. Тем не менее, статистические характеристики этого уравнения, являясь средними величинами, не отражают специфических отклонений и ошибок оценивания из-за различных уровней структурного сходства. По этой причине мы разработали каскадный расчет токсичности, в котором АМТ модели появляются последовательно.

На самом высоком уровне А (табл. 6) граница структурного сходства $T \geq 0.80$ и число вовлеченных в расчет структурных пар $NP=2$. Этому уровню соответствует 491 соединение (4.79% от общего числа соединений). При этом токсичность этих соединений может быть оценена со стандартным отклонением $s=0.33$. Следующий уровень В включает в себя соединения с $T \geq 0.70$ и $NP=2$, исключая соединения уровня А. Для таких 832 соединений (8.12% от общего числа соединений) стандартное отклонение равно 0.39. Подобным образом были получены модели для последующих уровней с исключением соединений предыдущего уровня.

Использование каскадной модели АМТ (уровни А-Е) позволило оценить острую токсичность 7085 соединений (около 69% от анализируемого ряда данных) по отношению к мышам со стандартным отклонением, сравнимым с экспериментальной ошибкой определения токсичности. Для оставшихся 3142

соединений (уровень F) стандартное отклонение равно 0.64, что эквивалентно приблизительно одной трети интервала токсичности для 97% изученных соединений. Поэтому это лучше, чем предсказание токсичности таких соединений качественным образом (низкая, средняя или высокая токсичность).

Таблица 6

Модель АМТ: параметры и статистические характеристики уравнения $\log(1/LD_{50})_{\text{эксп}} = a_0 + a_1 \log(1/LD_{50})_{\text{расч}}$ при различных уровнях структурного сходства (T_c) и числе структурных пар $NP=2$

Уровень	T_c	a_0	a_1	n	r^2	s	F	q^2
A	≥ 0.80	$0.03(\pm 0.02)$	$1.00(\pm 0.03)$	491	0.762	0.33	1567	0.756
B	≥ 0.70	$0.04(\pm 0.02)$	$0.96(\pm 0.02)$	832	0.673	0.39	1708	0.671
C	≥ 0.60	$0.01(\pm 0.02)$	$0.97(\pm 0.02)$	1291	0.673	0.38	2654	0.671
D	≥ 0.50	$0.04(\pm 0.01)$	$0.94(\pm 0.02)$	2010	0.581	0.44	2785	0.580
E	≥ 0.40	$0.10(\pm 0.01)$	$0.82(\pm 0.02)$	2461	0.442	0.48	1947	0.440
F	≥ 0.00	$0.16(\pm 0.01)$	$0.83(\pm 0.02)$	3142	0.337	0.64	1593	0.335

Для валидации модели исследуемый ряд из 10241 соединений был разделен на обучающую (8193 соединения) и тестовую (2048 соединений) выборки. В результате было установлено, что статистические критерии полученных уравнений для обучающей и тестовой выборок близки к критериям общей модели. В частности, величины s составили $0.26 \div 0.35$, 0.40 , $0.37 \div 0.40$, $0.43 \div 0.48$, $0.46 \div 0.52$, $0.60 \div 0.65$ для уровней A, B, C, D E, F соответственно.

Модель ЛРМПК

Модель локальных регрессионных моделей в перекрывающихся кластерах (ЛРМПК) стала результатом дальнейшего развития концепции дискретно-непрерывных QSAR моделей. Для ее тестирования был использован тот же ряд из 10241 соединения, дескрипторы и мера структурного сходства, что и в модели АМТ. Алгоритм ЛРМПК может быть представлен следующим образом.

1. Для рассматриваемого (i -го) соединения из выборки, содержащей n соединений, рассчитываются меры структурного сходства (T_c) с остальными $n-1$ соединениями. Полученные величины ранжируются по убыванию. Процедура повторяется для каждого соединения. В результате получается n отдельных

рядов соединений. 2. Фиксируется минимальный порог структурного сходства T_c^* , и в каждом из n рядов отбираются структурно-родственные соединения, имеющие T_c выше фиксированного уровня. Таким образом, формируются n перекрывающихся кластеров. 3. Для всех анализируемых соединений рассчитываются дескрипторы. 4. Для каждого кластера производится отбор дескрипторов на основе парных коэффициентов линейной корреляции и генерируются отдельные локальные регрессионные QSAR модели с использованием полного перебора всех возможных комбинаций из одного ($M=1$), двух ($M=2$) или трех ($M=3$) дескрипторов. Выбирается лучшее уравнение на основе критерия Фишера (F). В работе использовались линейные регрессионные модели при наличии в кластере ($5 M + 1$) и более структурных соседей. Это обеспечивает наличие минимум 5 соединений на один дескриптор в процессе перекрестного контроля с исключением по одному. 5. Полученные QSAR уравнения используются для расчета активности рассматриваемых соединений.

Из данных табл. 7 следует, что при задании в качестве граничного значения $T_c \geq 0.3$ и с учетом необходимого минимального числа структурных соседей, было получено 7759 кластеров, для которых было сконструировано 6425 однопараметровых, 1176 двухпараметровых и 158 трехпараметровых регрессионных уравнений для расчета токсичности выбранных соединений. Для оставшихся 2482 соединений не было обнаружено достаточного числа структурных соседей, необходимого для оценки токсичности с помощью описанного метода. В полученном уравнении свободный член и угловой коэффициент оказались близкими к идеальным (0.01 ± 0.01 и 0.99 ± 0.01 соответственно), а стандартное отклонение (0.51) - на уровне ошибки экспериментального измерения этого свойства для мышей разного возраста, пола с учетом условий их содержания и методики измерений.

Таблица 7

Модель ЛРМПК: параметры, статистические характеристики уравнения $\log(1/LD_{50})_{\text{экс}} = a_0 + a_1 \log(1/LD_{50})_{\text{расч}}$ при различных уровнях структурного сходства (T_c) и распределение числа моделей при использовании одного ($M=1$), двух ($M=2$) и трех ($M=3$) дескрипторов

№	T_c	a_0	a_1	n	r^2	s	F	q^2	M=1	M=2	M=3
1	≥ 0.3	$0.01(\pm 0.01)$	$0.99(\pm 0.01)$	7759	0.434	0.51	5958	0.434	6425	1176	158
2	≥ 0.4	$0.06(\pm 0.01)$	$0.91(\pm 0.01)$	5011	0.530	0.46	5646	0.529	4074	810	127
3	≥ 0.5	$0.06(\pm 0.01)$	$0.90(\pm 0.01)$	3157	0.621	0.40	5170	0.620	2514	540	103
4	≥ 0.6	$0.09(\pm 0.01)$	$0.87(\pm 0.02)$	1748	0.621	0.40	2862	0.618	1457	260	31
5	≥ 0.7	$0.15(\pm 0.02)$	$0.76(\pm 0.02)$	924	0.618	0.38	1490	0.613	814	99	11
6	≥ 0.8	$0.10(\pm 0.03)$	$0.88(\pm 0.03)$	346	0.659	0.37	666	0.643	272	74	0

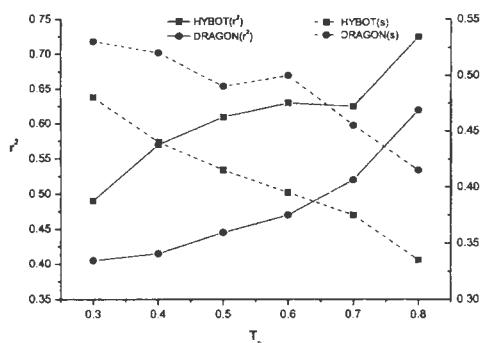


Рис. 2. Зависимость значений коэффициента корреляции (r^2) и стандартного отклонения (s) моделей ЛРМПК от величины порога структурного сходства (T_c) при использовании НУВОТ и DRAGON дескрипторов

Для сравнения были проведены расчеты токсичности на основе дескрипторов, полученных с помощью компьютерной программы DRAGON. Очевидно, что порог структурного сходства имеет существенное влияние на точность расчетов (рис. 2). Статистические критерии уравнений становятся лучше при большем значении данного порога. Но одновременно количество получаемых уравнений, а, следовательно, и число соединений с рассчитанным

значением токсичности уменьшается. Это явно демонстрирует эффективность использования концепции структурного сходства для оценки свойств/активности соединений. Также очевидно, что лучшие результаты расчетов получаются при использовании дескрипторов программы HUBOT.

Валидация модели ЛРМПК выполнена с использованием обучающей и тестовой выборок (см. модель АМТ). Близость статистических критериев уравнений, полученных на основе изученных рядов соединений, свидетельствует об устойчивости анализируемой модели.

Глава 5. Теоретические разработки в области исследования количественной связи «структура – активность» и «структура – свойство»

Рекуррентная модель

На основе рекуррентных уравнений проведено исследование острой токсичности 85 органических соединений из 7 гомологических рядов по отношению к 6 организмам: бактерии *Vibrio fisheri*, инфузории *Tetrahymena pyriformis*, рыбам *Poecilia reticulata* и *Pimephales promelas*, мышам и крысам (внутривенно). Токсичность менялась в широких пределах: для водных организмов от $\log(1/LC_{50})=-2.96$ до $\log(1/EC_{50})=3.69$, для теплокровных организмов от $\log(1/LD_{50})=-2.17$, до $\log(1/LD_{50})=1.09$.

Зависимость токсичности от числа атомов углерода во всех исследованных гомологических рядах имеет прямолинейный характер или стремится к некоторому пределу. Поэтому для аппроксимации данных может быть использовано простое рекуррентное уравнение вида:

$$T_n = A + B T_{n-1} \quad (20)$$

где T_n – токсичность текущего n -го члена гомологического ряда, T_{n-1} – токсичность предшествующего члена гомологического ряда, A и B – параметры. Полученные уравнения характеризуются высокими статистическими показателями. Например, в большинстве случаев, $r^2 > 0.9$. Часть представленных зависимостей имеет предельное значение токсичности ($B < 1$), другая часть не имеет предела ($B > 1$). При этом коэффициент A варьирует от 0.13 до 0.67.

Полезность математических моделей для описания токсичности можно связать с качеством экстраполяции данных. С этой целью мы сопоставили три модели: рекуррентную, линейную и параболическую. В двух последних моделях в качестве независимой переменной фигурировало число метиленовых групп в гомологическом ряду. Для сравнения были выбраны данные по токсичности с учетом количества молекул в ряду и с максимальным представлением всех использованных типов химических соединений и биологических объектов. В результате было отобрано 5 комбинаций гомологический ряд/биологический объект. Устойчивость и предсказательная способность моделей была оценена с помощью кросс-валидации с отбрасыванием 3-х последних точек и расчетом стандартных отклонений s . При этом было установлено, что для линейной модели значение $s=0.36$, для параболической модели $s=0.49$ и для рекуррентной модели $s=0.33$.

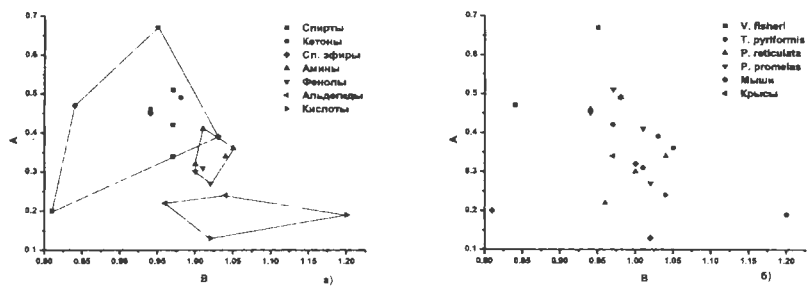


Рис. 3. Параметры рекуррентных уравнений на плоскости А-В с учетом типа гомологического ряда (а) и вида биологического объекта (б)

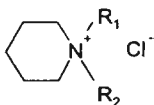
Рекуррентные модели могут быть выражены в виде точек на плоскости А-В с учетом типа гомологического ряда и вида биологического объекта. На рис. За отчетливо видны три кластера. В первый кластер входят спирты, кетоны и сложные эфиры. Во второй кластер – амины и фенолы. Третий кластер состоит из альдегидов и кислот. Таким образом, представленный метод графического

анализа позволяет различать гомологические ряды между собой, относя их к различным типам токсического действия.

Для анализа токсикологической близости биообъектов между собой мы использовали расстояние между точками (рис. 3б). Обращает на себя внимание уникальность данных для бактерии *V. fischeri* ввиду отсутствия в ближайших окрестностях аналогичных моделей токсичности. Близость параметров рекуррентных уравнений, описывающих токсичность химических соединений по отношению к рыбам *P. reticulata* и *P. promelas* не вызывает удивления: все-таки это очень сходные организмы. Из наших данных также следует сходство моделей токсичности для водных организмов *T. pyriformis* и *P. reticulata*, а также *T. pyriformis* и *P. promelas*.

Ретардантная активность солей четвертичного аммония

Для исследования антигиббереллиновой (ретардантной) активности был использован ряд из 12 хлоридов N,N-диалкилпиперидиниев:



где $R_1 = \text{CH}_3, \text{C}_2\text{H}_5, \text{C}_3\text{H}_7, \text{C}_4\text{H}_9, \text{CH}_2\text{C}_6\text{H}_5$; $R_2 = \text{CH}_3, \text{C}_4\text{H}_9, \text{C}_6\text{H}_5, \text{CH}_2\text{CH}_2\text{OCH}_3, \text{CH}_2\text{CH}_2\text{OC}_2\text{H}_5, \text{CH}_2\text{CH}_2\text{OCH}_2\text{C}_6\text{H}_5$. При этом антигиббереллиновая активность (IC_{50} , моль/л) менялась в пределах от $\log(1/\text{IC}_{50})=2.98$ до $\log(1/\text{IC}_{50})=5.95$; интервалы изменения молекулярной поляризуемости (α , \AA^3), максимального свободноэнергетического протонноакцепторного фактора (C_a^{max}) и стерического параметра ($-E_s$) составили $14.58 \div 38.21, 0.00 \div 1.49, 7.09 \div 10.25$ соответственно. В результате QSAR моделирования была получена следующая регрессионная модель:

$$\log(1/\text{IC}_{50}) = 14.93(\pm 1.82) + 0.14(\pm 0.04) \alpha - 0.83(\pm 0.19) C_a^{\text{max}} + 1.52(\pm 0.30) E_s \quad (21)$$

$n=12; r^2=0.895; s=0.38; q^2=0.788$

Из представленных данных ясно, что существенную роль в формировании антигиббереллиновой активности исследованных солей играют поляризуемость и протонноакцепторная способность молекул, а также

стерический параметр, причем наибольший вклад вносит именно он. Определяющее влияние стерического фактора связано, по-видимому, с тем, что четвертичный атом азота встроен в пиперидиновое кольцо. Это ограничивает конформационные возможности молекулы при ее взаимодействии с активным центром рецептора и предъявляет повышенные требования к молекулярным свойствам, которые привносят фрагменты R_1 и R_2 .

Фрактальная размерность гистограмм межатомных расстояний

Для описания геометрической структуры соединений были использованы гистограммы (дискретные функции распределения) межатомных расстояний, которые представлялись в виде плоских диаграмм, состоящих из набора закрашенных прямоугольников (баров). Ось абсцисс разбивалась с определенным шагом (разрешением) ΔR на интервалы, количество которых зависело от значений межатомных расстояний R , меняющихся от минимального значения R_{\min} до максимального значения R_{\max} . На оси ординат откладывалась частота f , т.е. количество межатомных расстояний, попадающих в заданный интервал, и варьирующая в виде целых чисел от 0 до f_{\max} . Эта ось также разбивалась на интервалы с определенным шагом Δf . Полученные гистограммы оцифровывались путем их представления в виде квадратных бинарных матриц, где значение 1 соответствовало закрашенным участкам гистограммы, а 0 – всему остальному.

Фрактальная размерность определялась «клеточным» методом с помощью компьютерной программы, путем покрытия оцифрованного изображения клетками переменной величины и подсчетом числа клеток, которые пересекают закрашенные области гистограмм. Локальная фрактальная размерность рассчитывалась по уравнению:

$$\log N(L) = \log C - D \log L \quad (22)$$

где: $N(L)$ – минимальное число клеток со стороной L , необходимых для покрытия фрактала; D – «клеточная» фрактальная размерность (со знаком -); C – константа. Пример расчета представлен на рис. 4. Подобным образом были проведены расчеты для 73 органических молекул, относящихся к 6 различным

классам: алканы, спирты, эфиры, кетоны, амины, ароматические соединения. При этом число атомов N_a меняется от 5 до 50, а максимальная длина межатомных расстояний R_{max} варьирует в пределах от 1.82 Å до 20.57 Å. Минимальная величина фрактальной размерности D составляет 1.00 (метан), а максимальная – 1.58 (1,2:5,6-дибензантрацен).

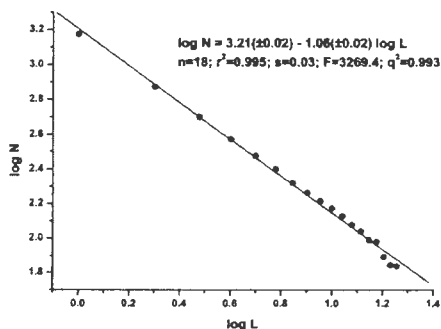


Рис. 4. Зависимость числа клеток (N) от их размера (L) (уравнение (22)) для молекулы метанола

Анализ полученных данных свидетельствует о существовании тенденции между ростом числа атомов в молекуле и фрактальной размерностью. Однако, наблюдающиеся значительные отклонения от тренда, говорят о том, что число атомов не является единственным фактором. Например, четыре изомерных спирта с общей брутто-формулой $C_4H_{10}O$: 1-бутанол(I), 2-метил-1-пропанол(II), 2-бутанол(III) и 2-метил-2-пропанол(IV) имеют по 15 атомов, но различаются по фрактальной размерности: D при этом меняется от 1.20 до 1.38. Наблюдающееся явление может быть объяснено, если принять во внимание точечные группы симметрии молекул. Действительно, соединения (I) и (IV) относятся к группе C_s и имеют $D=1.25$ и 1.20 соответственно. Напротив, соединения (II) и (III) принадлежат к группе C_1 и их фрактальные размерности равны 1.33 и 1.38 соответственно. Количественной мерой, отражающей количество атомов в молекуле и принадлежность молекул к той или иной точечной группе симметрии, в первом приближении, может являться число

баров в гистограммах межатомных расстояний N_b . С учетом этого для соединений с различными точечными группами симметрии были сконструированы следующие модели:

$$C_{2v}: D = 0.97(\pm 0.03) + 0.81(\pm 0.06) \log(N_b/N_a) \quad (23)$$

$$n=14; r^2=0.935; s=0.04; F=172; q^2=0.889$$

$$C_s: D = 0.91(\pm 0.03) + 0.70(\pm 0.06) \log(N_b/N_a) \quad (24)$$

$$n=23; r^2=0.851; s=0.04; F=120; q^2=0.810$$

$$C_1: D = 0.91(\pm 0.06) + 0.63(\pm 0.08) \log(N_b/N_a) \quad (25)$$

$$n=18; r^2=0.785; s=0.03; F=58.5; q^2=0.682$$

Уравнения (23)-(25) свидетельствуют о существовании корреляционной связи между фрактальной размерностью и нормированным числом баров. Эта связь становится понятной, если проанализировать физический смысл величины N_b/N_a . По сути, она представляет собой число баров, которое приходится на один атом, т.е. характеризует плотность заполнения пространства гистограммы. Поэтому фрактальная размерность, являясь количественной мерой степени изрезанности или сложности пространственной структуры молекул (в пределе, крайняя изрезанность – это полностью заполненное пространство), оказывается в статистической связи с величиной N_b/N_a .

Полезность нового предлагаемого дескриптора может быть продемонстрирована на примере установления количественных соотношений между фрактальной размерностью и плотностью для гомологического ряда алканов (пентан, гексан,...эйкозан, пентакозан, триаконтан), которые при нормальных условиях представляют собой жидкости или твердые тела:

$$d_4^{20} (\text{г/см}^3) = -0.21(\pm 0.05) + 0.70(\pm 0.03) D \quad (26)$$

$$n=18; r^2=0.965; s=0.01; F=442; q^2=0.956$$

Полученное регрессионное уравнение свидетельствует о существовании тесной корреляционной связи между соответствующими микро- и макромолекулярными характеристиками. В изученном ряду соединений возрастание сложности молекулярной структуры сопровождается увеличением плотности вещества.

Линейная динамическая модель

Для описания структуры химических соединений была использована функция радиального распределения в приближении жесткой модели молекулы (отсутствие внутримолекулярных колебаний):

$$f(R) = \frac{2}{\pi} \int_0^{s_{\max}} s M(s) \exp(-bs^2) \sin sR ds \quad (27)$$

здесь $f(R)$ - функция радиального распределения, R - межатомное расстояние, $M(s)$ - функция молекулярного рассеяния, s - угловой параметр.

В качестве моделируемого свойства выступала обратная функция давления насыщенного пара от температуры $T=f(P)$. Количественную связь между исследуемыми кривыми устанавливали на основе линейной динамической модели с одним входом, одним выходом и импульсной переходной функцией. Непосредственно для расчетов использовалось уравнение:

$$Y(f) = H(f) X(f) \quad (28)$$

здесь $X(f)$ и $Y(f)$ – преобразования Фурье для входной и выходной кривой соответственно, а $H(f)$ - частотная характеристика. Частотную характеристику вычисляли путем усреднения по всем кривым на основе соотношения для взаимного спектра входного и выходного процессов:

$$S_{xy}(f) = H(f) S_{xx}(f) \quad (29)$$

здесь $S_{xy}(f)$ – взаимная двусторонняя спектральная плотность, $S_{xx}(f)$ – двусторонняя спектральная плотность входной кривой.

Качество исследуемой модели оценивали с помощью функции когерентности, изменяющейся в пределах от 0 до 1 и вычисляемой по формуле:

$$\gamma^2(f) = \frac{|S_{xy}(f)|^2}{S_{xx}(f) S_{yy}(f)} \quad (30)$$

где $S_{yy}(f)$ – двусторонняя спектральная плотность выходной кривой. В дальнейшем для сравнения различных моделей использовали среднее значение γ^2 , полученное путем усреднения функции когерентности по всем частотам. Расчеты выполняли с использованием модифицированной компьютерной программы на основе быстрого преобразования Фурье. В качестве входных

и/или выходных процессов выступали $f(R)$ и обратная функция давления насыщенного пара от температуры $T=f(P)$. Оцифровку зависимостей проводили путем дискретизации независимых переменных с использованием 16, 32, 64, 128, 256 точек (уровни дискретизации m) и вычислением соответствующих значений функций.

Учитывая разный физический смысл входных и выходных кривых, для получения статистически значимых корреляционных зависимостей использовали модификацию линейной динамической модели. Для этого на вход системы подавали участки кривых различной длины. При использовании $f(R)$ в качестве входной кривой (модель «структура – активность») проводили варьирование ее длины путем деления интервала R (0 – 16 Å) на 160 частей и парной комбинации полученных значений R_{\min} и R_{\max} . При этом длина выходной кривой $T=f(P)$ оставалась неизменной и была максимальной ($P_{\min}=10$ мм Hg, $P_{\max}=1500$ мм Hg). Аналогичный подход использовали для входной кривой $T=f(P)$ (модель «активность – структура»). В этом случае интервал P (10 – 1500 мм Hg) делили на 50 частей и комбинировали величины P_{\min} и P_{\max} . Длина кривой $f(R)$ была постоянной ($R_{\min}=0$ Å, $R_{\max}=16$ Å). Полученные модели сортировали по величине остаточного стандартного отклонения и отбирали те из них, которые имели минимальное значение. Для исследования был использован ряд из 10 н-алканов (пропан, бутан,...додекан).

В табл. 8 приведены характеристики лучших моделей типа «структура – активность», полученных путем перебора 12880 реализаций различной длины на каждом уровне дискретизации. Учитывая, что модель 1 имеет минимальную величину $s=5.0$ и максимальное значение $\gamma^2=0.988$, она была выбрана в качестве финальной модели типа «структура-активность». С использованием этой модели были рассчитаны функции $T=f(P)$ для 10 исследуемых соединений:

$$T_{\text{расч}} = -3.5(\pm 1.7) + 1.01(\pm 0.01) T_{\text{эксп}} \quad (31)$$

$n=160$; $r^2=0.997$; $s=5.0$; $q^2=0.996$

Уравнение (31) свидетельствует о том, что модель 1 адекватно описывает кривую $T=f(P)$.

Таблица 8

Порядковый номер модели (№), уровни дискретизации (m), минимальная ($R_{\text{мин}}$) и максимальная ($R_{\text{макс}}$) величина межатомного расстояния, остаточное стандартное отклонение (s) и среднее значение функции когерентности (γ^2) для моделей «структура – активность»

№	m	$R_{\text{мин}}, \text{\AA}$	$R_{\text{макс}}, \text{\AA}$	s	γ^2
1	16	1.1	13.9	5.0	0.988
2	32	0.8	12.7	5.4	0.944
3	64	1.4	15.3	5.2	0.976
4	128	0.0	15.5	6.6	0.956
5	256	0.0	15.6	6.6	0.933

В рамках рассматриваемой динамической модели также может решаться «обратная» задача QSAR, т.е. предсказание структур с заданными свойствами. С этой целью мы сконструировали специальную модель «активность – структура», где в качестве входа используется зависимость $T=f(P)$ переменной длины, а в качестве выхода – полноразмерная $f(R)$. Учитывая, что точность описания функции радиального распределения по оси ординат (интенсивность) на уровне величины 0.06 представляется приемлемой, а максимальное разрешение по оси абсцисс (межатомное расстояние) достигается при $m=256$ и составляет 0.06 Å, в качестве рабочей модели «активность – структура» была использована модель 10 (табл. 9). Используя эту модель и схему кросс-валидации с выбором по одному, были рассчитаны величины $f(R)$ для всех исследованных n-алканов. Ниже представлено соответствующее регрессионное уравнение:

$$f(R)_{\text{расч}} = -0.001(\pm 0.001) + 1.00(\pm 0.01) f(R)_{\text{экср}} \quad (32)$$

$$n=2560; r^2=0.971; s=0.061; q^2=0.966$$

Параметры и коэффициенты регрессионного уравнения (32) свидетельствуют о высоком качестве модели 10, которая может быть применена для вычисления $f(R)$ в исследованном ряду n-алканов.

Таблица 9

Порядковый номер модели (№), уровни дискретизации (m), минимальная ($P_{\text{мин}}$) и максимальная ($P_{\text{макс}}$) величина давления насыщенного пара, остаточное стандартное отклонение (s) и среднее значение функции когерентности (γ^2) для моделей «активность - структура»

№	m	$P_{\text{мин}}$, мм Hg	$P_{\text{макс}}$, мм Hg	s	γ^2
6	16	1470.2	1500.0	0.018	0.992
7	32	10.0	39.8	0.047	0.893
8	64	1470.2	1500.0	0.072	0.919
9	128	1470.2	1500.0	0.059	0.943
10	256	1470.2	1500.0	0.061	0.940

Таким образом, задача предсказания структур с заданными свойствами в рамках модифицированной линейной динамической модели может быть решена путем конструирования модели типа «активность – структура», расчета на ее основе функции радиального распределения для молекулы с заданными свойствами и сравнения с $f(R)$ модельного соединения с известной трехмерной структурой.

ВЫВОДЫ

1. С использованием аддитивно-мультипликативного подхода и термодинамических характеристик Н-комплексов созданы энтальпийная и свободноэнергетическая шкалы водородной связи и рассчитаны дескрипторы, характеризующие способность соединений к образованию Н-связи (факторы водородной связи).
2. Разработана количественная модель для описания липофильности органических соединений, выраженной в виде коэффициента распределения в системе н-октанол – вода, на основе поляризуемости и свободноэнергетических протонакцепторных факторов.
3. На базе стандартных и оригинальных подходов сконструированы и протестированы модели острой токсичности органических соединений по отношению к водным организмам и грызунам с применением факторов

водородной связи и других дескрипторов. Продемонстрирована их применимость для предсказания токсичности больших массивов органических соединений разнообразных химических классов.

4. Разработана рекуррентная модель токсичности в гомологических рядах соединений. Выявлено, что с помощью этой модели может быть установлен тип токсического действия ряда и токсикологическая близость биообъектов между собой.

5. С использованием физико-химических дескрипторов: молекулярной поляризуемости, свободноэнергетического протоноакцепторного фактора и стерического параметра разработаны модели антигипбереллиновой (ретардантной) активности солей четвертичного аммония.

6. Для описания структуры химических соединений создан и применен новый дескриптор: фрактальная размерность. Установлено, что предлагаемый дескриптор описывает сложность геометрической структуры молекул и может быть использован при исследовании связи «структура – свойство» и «структура – активность».

7. Разработан новый подход для установления количественной связи между структурой и свойствами (активностью) химических соединений на основе модифицированной линейной динамической модели, который может быть использован в качестве рабочего инструмента при проведении QSAR/QSPR исследований.

8. На основе теоретических разработок получен ряд практических результатов: создана база данных по термодинамическим характеристикам комплексов с водородной связью; сформирована база данных, содержащая Н-донорные и Н-акцепторные факторы органических соединений; разработаны компьютерные программы, ориентированные на использование в QSAR/QSPR исследованиях.

Основные публикации по теме диссертации

1. Раевский О.А., Григорьев В.Ю., Соловьев В.П. Оценка электронодонорных и акцепторных функций ионизированных атомов и групп в биологически

- активных веществах на основе термодинамических данных // Хим.-фармац. журн. 1984. Т. 18. № 5. С. 578-582.
2. Раевский О.А., Григорьев В.Ю. Сборник задач по курсу «Конструирование биологически активных веществ». М.: МХТИ, 1986. 80 с.
 3. Раевский О.А., Григорьев В.Ю., Соловьев В.П., Иванов А.Н., Соколов В.Б., Мартынов И.В. Электронодонорные и акцепторные функции физиологически активных и модельных соединений. X. Электронодонорные функции О-метилхлорформимино-О-этилметилфосфоната // Журн. общ. химии. 1987. Т. 57. № 9. С. 2073-2078.
 4. Григорьев В.Ю., Соловьев В.П., Раевский О.А., Иванов А.Н., Соколов В.Б., Чепакowa Л.А., Брель В.К., Мартынов И.В. Электронодонорные и акцепторные функции физиологически активных и модельных соединений. XII. Термодинамические параметры молекулярных комплексов некоторых фосфорилированных оксимов с фенолом и дифениламиноm // Журн. общ. химии. 1988. Т. 58. № 4. С. 761-766.
 5. Раевский О.А., Григорьев В.Ю., Соловьев В.П., Мартынов И.В. Электронакцепторные энтальпийные факторы фенолов // Докл. АН СССР. 1988. Т. 298. № 5. С. 1166-1169.
 6. Раевский О.А., Григорьев В.Ю., Соловьев В.П., Мартынов И.В. Электронакцепторные энтропийные факторы фенолов // Докл. АН СССР. 1988. Т. 299. № 6. С. 1433-1438.
 7. Раевский О.А., Соловьев В.П., Григорьев В.Ю. Термодинамические характеристики водородной связи фенолов с органическими основаниями // Рукопись деп. ВИНТИ 1988. № 1001-В88. 83 с.
 8. Raevsky O.A., Solov'ev V.P., Solotnov A.F., Grigor'ev V.Yu., Kireev D.B. The Electron-Donor Function of the Phosphorylic Oxygen // Proceedings of the 11-th International Conference on Phosphorus Chemistry. Tallinn. 1989. P. I-75.
 9. Раевский О.А., Григорьев В.Ю., Соловьев В.П. Моделирование связи структура – активность. II. Оценка электронодонорных и акцепторных

- функций активных центров в молекулах физиологически активных веществ // Хим.-фармац. журн. 1989. Т. 23. № 11. С. 1294-1300.
10. Григорьев В.Ю., Соловьев В.П., Додонов М.В., Брель В.К., Мартынов И.В., Раевский О.А. Электронодонорные и акцепторные функции физиологически активных и модельных соединений. XIV. Электроноакцепторная способность группы N-H дифторнитроацетанилидов // Журн. общ. химии. 1989. Т. 59. № 4. С. 846-851.
11. Григорьев В.Ю., Соловьев В.П., Раевский О.А., Иванов А.Н., Соколов В.Б., Мартынов И.В. Электронодонорные и акцепторные функции физиологически активных и модельных соединений. XV. Использование мультипликативного правила факторов для оценки термодинамических характеристик водородной связи дифторнитроацетанилидов с фосфорилированными оксимами // Журн. общ. химии. 1990. Т. 60. № 3. С. 527-530.
12. Гафуров Р.Г., Чистяков В.Г., Кокурин А.В., Григорьев В.Ю., Прошин А.Н., Раевский О.А., Муромцев Г.С., Мартынов И.В. Связь химической структуры и активности регуляторов роста растений. Сообщение 1. Влияние стерического параметра катиона на антигипбереллиновую активность хлоридов диалкилпиперидиниев // Агрохимия. 1991. № 12. С. 81-85.
13. Raevsky O.A., Grigor'ev V.Yu., Solov'ev V.P., Kireev D.B., Sapegin A.M., Zefirov N.S. Drug Design H-bonding Scale // In: QSAR: Rational Approaches to the Design of Bioactive Compounds / Silipo C., A. Vittoria A., Eds. Amsterdam: Elsevier Science Publishers, 1991. P. 135-138.
14. Григорьев В.Ю., Раевская О.Е., Раевский О.А. База данных «Термодинамика водородной связи» // Тезисы IX Всесоюзной конференции «Химическая информатика». Черноголовка, МО. 1992. С. 33.
15. Raevsky O.A., Grigor'ev V.Yu., Kireev D.B., Lukoyanov N.V., Glinka R., Brzezinska E. QSAR Models of β - Adrenergic Activity of Some Substituted Phenethylamines // Proceedings of the 12-th International Symposium on Medicinal Chemistry. Basel. 1992. P. P-037A.

16. Raevsky O.A., Grigor'ev V.Yu., Mednikova E. QSAR H-Bonding Descriptors // Proceedings of the 9-th European Symposium on Structure-Activity Relationships: QSAR and Molecular Modelling. Strasbourg. 1992. P. O-22.
17. Raevsky O.A., Grigor'ev V.Yu., Kireev D.B., Zefirov N.S. Correlation Analysis and H-bond Ability in Framework of QSAR // J. Chim. Phys. 1992. Vol. 89. P. 1747-1753.
18. Raevsky O.A., Grigor'ev V.Yu., Kireev D.B., Zefirov N.S. Complete Thermodynamic Description of H-Bonding in the Framework of Multiplicative Approach // Quant. Struct.-Act. Relat. 1992. Vol. 11. № 1. P. 49-63.
19. Raevsky O., Dolmatova L., Grigor'ev V., Lisyansky I., Bondarev S. Molecular Recognition Descriptors in QSAR // In: QSAR and Molecular Modelling: Concepts, Computational Tools and Biological Applications / Sanz F., Giraldo J., Eds. Barselona: Prous Science Publ., 1995. P. 241-245.
20. Раевский О.А., Григорьев В.Ю., Трепалин С.В., Шапер К. Расчет гидрофобности на основе поляризуемости и водородной связи // Тезисы IV Российского конгресса «Человек и лекарство». Москва. 1997. С. 288.
21. Раевский О.А., Григорьев В.Ю., Трепалин С.В., Герасименко В.А., Раздольский А.Н., Шапер К. Оптимизация компьютерного конструирования лекарств на основе физико-химических параметров // Тезисы VI Российского конгресса «Человек и лекарство». Москва. 1999. С. 463.
22. Раевский О.А., Григорьев В.Ю. Количественное описание липофильности органических соединений на основе поляризуемости и акцепторной способности к образованию водородной связи // Хим.-фармац. журн. 1999. Т. 33. № 5. С. 46-49.
23. Раевский О.А., Григорьев В.Ю., Трепалин С.В. Свидетельство об официальной регистрации программы для ЭВМ HVBOT (Hydrogen Bond Thermodynamics) № 990090 от 26 февраля 1999 г., Москва, Федеральная служба по интеллектуальной собственности, патентам и товарным знакам.

24. Григорьев В.Ю. Расчет вклада водородной связи в энтальпию испарения и нормальную точку кипения органических соединений // Рукопись деп. ВИНТИ 04.06.99. № 1820-B99. 22 с.
25. Григорьев В.Ю. Комплексный метод оценки протонодонорной способности фенолов // Рукопись деп. ВИНТИ 19.07.2000. № 2012-B00. 13 с.
26. Раевский О.А., Скворцов В.С., Григорьев В.Ю., Трепалин С.В. Свидетельство об официальной регистрации программы для ЭВМ Программный комплекс НУВОТ (HYdrogen BOnd Thermodynamics) in UNIX № 2002610496 от 4 апреля 2002 г., Москва, Федеральная служба по интеллектуальной собственности, патентам и товарным знакам.
27. Гафуров Р.Г., Григорьев В.Ю., Прошин А.Н., Чистяков В.Г., Мартынов И.В., Зефилов Н.С. Роль молекулярных параметров солей четвертичного аммония в формировании их антигипбереллиновой (ретардантной) активности // Биоорг. химия. 2004. Т. 30. № 6. С. 656-662.
28. Гафуров Р.Г., Григорьев В.Ю., Прошин А.Н., Чистяков В.Г., Мартынов И.В., Зефилов Н.С. Роль молекулярных параметров солей четвертичного аммония в формировании их антигипбереллиновой (ретардантной) активности // Докл. РАН. 2004. Т. 394. № 5. С. 710-714.
29. Raevsky O.A., Grigor'ev V.Yu., Raevskaja O.E., Schaper K.-J. Physicochemical Properties/Descriptors Governing the Solubility and Partitioning of Chemicals in Water-Solvent-Gas systems. Part.1. Partitioning between Octanol and Air // SAR QSAR Environ. Res. 2006. Vol. 17. № 3. P. 285-297.
30. Grigor'ev V.Yu., Weber E.E., Raevsky O.A. Classification SAR Models of Chemicals Toxicity to Guppy, Fathead Minnow and Rainbow Trout on the Basis of Structure Similarity // Fourth International Symposium on Computational Methods in Toxicology and Pharmacology Integrating Internet Resources. Moscow. 2007. P. 99.
31. Tikhonova O.V., Skvortsov V.S., Grigor'ev V.Yu., Raevsky O.A. 3D-QSAR Models of Irreversible and Reversible Inhibitors of Acetylcholinesterase // Fourth

- International Symposium on Computational Methods in Toxicology and Pharmacology Integrating Internet Resources. Moscow. 2007. P. 162.
32. Raevsky O.A., Grigor'ev V.Yu., Weber E.E., Dearden J.C. Classification and Quantification of the Toxicity of Chemicals to Guppy, Fathead Minnow and Rainbow Trout. Part 1. Nonpolar Narcosis Mode of Action // QSAR Comb. Sci. 2008. Vol. 27. № 11-12. P. 1274 – 1281.
33. Raevsky O.A., Grigor'ev V.Yu., Dearden J.C., Weber E.E. Classification and Quantification of the Toxicity of Chemicals to Guppy, Fathead Minnow, and Rainbow Trout. Part 2. Polar Narcosis Mode of Action // QSAR Comb. Sci. 2009. Vol. 28. № 2. P. 163-174.
34. Раевский О.А., Григорьев В.Ю., Тихонова О.В. Развитие моделей взаимосвязи структуры и токсического действия химических соединений по отношению к Guppy // Хим.-фармац. журн. 2009. Т. 43. № 3. С. 100-104.
35. Raevsky O.A., Grigor'ev V.Yu., Modina E.A., Raevskaya O.E., Liplavsky Ja.V. Classification- and Regression-Based QSAR of Chemical Toxicity on the Basis of Structural and Physicochemical Similarity // Workshop on "Satisfying REACH Requirements in Predictive Toxicology". Rome. 2009. P. 12-13.
36. Григорьев В.Ю., Раевский О.А. Рекуррентная модель острой токсичности гомологического ряда нормальных спиртов // Материалы XVII Российского национального конгресса «Человек и лекарство». Москва. 2010. С. 600.
37. Раевский О.А., Григорьев В.Ю., Модина Е.А. Компьютерные модели взаимосвязи структуры органических соединений и их острой токсичности // Материалы XVII Российского национального конгресса «Человек и лекарство». Москва. 2010. С. 708-709.
38. Raevsky O.A., Grigor'ev V.Yu., Modina E.A., Worth A.P. Classification- and Regression-Based QSAR of Acute Chemical Rodent Toxicity // Proceedings of the 18-th European Symposium on Quantitative Structure-Activity Relationships. Rhodes. 2010. P. 77-78.

39. Raevsky O.A., Grigor'ev V.Yu., Modina E.A., Worth A.P. Prediction of Acute Toxicity to Mice by the Arithmetic Mean Toxicity (AMT) Modelling Approach // SAR QSAR Environ. Res. 2010. Vol. 21. № 3-4. P. 265-275.
40. Raevsky O.A., Grigor'ev V.Y., Liplavskaya E.A., Worth A.P. Prediction of Acute Rodent Toxicity on the Basis of Chemical Structure and Physicochemical Similarity // Mol. Inf. 2011. Vol. 30. № 2-3. P. 267-275.
41. Григорьев В.Ю., Раевский О.А. Фрактальная размерность гистограмм межатомных расстояний – новый 3D дескриптор молекулярной структуры // Журн. общ. химии. 2011. Т. 81. № 3. С. 353-359.
42. Григорьев В.Ю., Раевский О.А. Рекуррентная модель острой токсичности в гомологических рядах органических соединений // Хим.-фармац. журн. 2011. Т. 45. № 4. С. 124-129.
43. Григорьев В.Ю., Раевский О.А. Гистограмма межатомных расстояний – новый 3D дескриптор молекулярной структуры // Материалы XVIII Российского национального конгресса «Человек и лекарство». Москва. 2011. С. 552.
44. Липлавская Е.А., Григорьев В.Ю., Раевский О.А. Предсказание острой токсичности органических соединений по отношению к грызунам на основе структурного и физико-химического сходства // Материалы XVIII Российского национального конгресса «Человек и лекарство». Москва. 2011. С. 612.
45. Raevsky O.A., Grigoriev V.Yu., Liplavskaya E.A., Worth A.P. Calculations of Acute Intravenous Toxicity in Mice Based on Local Regression Models in Superoverlapping Clusters (LRMSC) // Biochemistry (Moscow) Supplement Series B: Biomedical Chemistry. 2011. Vol. 5. № 4. P. 346-356.
46. Григорьев В.Ю., Раевский О.А. Исследование количественной связи между функцией радиального распределения и кривой давления насыщенного пара в ряду n-алканов на основе модифицированной линейной динамической модели // Журн. физ. химии. 2012. Т. 86. № 3. С. 417-422.

Сдано в печать 05.07.12. Подписано в печать 06.07.12. Формат 60х90 1/16
Объем 2,75 п. л. Заказ 99. Тираж 100

Отпечатано в типографии ИПХФ РАН
142432, Московская обл., г. Черноголовка, пр-т ак. Семенова, 5
Тел.: 8(49652)2-19-38

102